Maths for ML

Nipun Batra

IIT Gandhinagar

July 31, 2025

Outline

Why Vector Notation?

Motivation

Matrix/vector notation makes ML math cleaner, more efficient, and easier to implement

Why Vector Notation?

Motivation

Matrix/vector notation makes ML math cleaner, more efficient, and easier to implement

Why Vector Notation?

Motivation

Matrix/vector notation makes ML math cleaner, more efficient, and easier to implement

Benefits:

- **Concise:** Write complex operations in one line

Why Vector Notation?

Motivation

Matrix/vector notation makes ML math cleaner, more efficient, and easier to implement

Benefits:

- **Concise:** Write complex operations in one line
- **Efficient:** Vectorized computations are faster

Why Vector Notation?

Motivation

Matrix/vector notation makes ML math cleaner, more efficient, and easier to implement

Benefits:

- **Concise:** Write complex operations in one line
- **Efficient:** Vectorized computations are faster
- **General:** Works for any number of data points/features

The Error Vector ϵ

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

The Error Vector $\boldsymbol{\epsilon}$

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

Column Vector (Error Vector)

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

The Error Vector $\boldsymbol{\epsilon}$

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

Column Vector (Error Vector)

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

The Error Vector $\boldsymbol{\epsilon}$

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

Column Vector (Error Vector)

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

Row Vector (Transpose)

$$\boldsymbol{\epsilon}^T = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_N \end{bmatrix}_{1\times N}$$

The Error Vector $\boldsymbol{\epsilon}$

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

Column Vector (Error Vector)

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

Row Vector (Transpose)

$$\boldsymbol{\epsilon}^T = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_N \end{bmatrix}_{1 \times N}$$

The Error Vector $\boldsymbol{\epsilon}$

Setup: For N data points, each has prediction error $\epsilon_i = y_i - \hat{y}_i$

Column Vector (Error Vector)

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

Row Vector (Transpose)

$$\boldsymbol{\epsilon}^T = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_N \end{bmatrix}_{1\times N}$$

Sum of Squared Errors (Matrix Form)

Scalar Property

1. For a scalar s
$$s = s^T$$

Gradient Vector

2.

$$\nabla_\theta s = \begin{bmatrix} \frac{\partial s}{\partial \theta_1} \\ \frac{\partial s}{\partial \theta_2} \\ \vdots \\ \frac{\partial s}{\partial \theta_N} \end{bmatrix}$$

Matrix Transpose and Multiplication

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Matrix Transpose and Multiplication

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Matrix Transpose and Multiplication

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$Z = X^T X = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}_{2\times 2}$$

Matrix Transpose and Multiplication

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$X^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$Z = X^T X = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}_{2 \times 2}$$

Z has a property $Z_{ij} = Z_{ji} \implies Z^T = Z$

Quadratic Form

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Quadratic Form

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z\theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Quadratic Form

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z \theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

Quadratic Form

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z\theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z\theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

$$\theta^T Z\theta = e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2$$

The term $\theta^T Z\theta$ is a scalar.

Matrix Rank

Since X has fewer rows than columns, its maximum rank is equal to the maximum number of linearly independent rows. And because neither row is linearly dependent on the other row, the matrix has 2 linearly independent rows; so its rank is 2.

Matrix Inverse

For a matrix A, its inverse A^{-1} satisfies $AA^{-1} = A^{-1}A = I_n$ where $\mathrm{I_n}$ is the identity matrix.

Matrix Inverse

For a matrix A, its inverse A^{-1} satisfies $AA^{-1} = A^{-1}A = I_n$ where $\mathrm{I_n}$ is the identity matrix.
Below, with an example, we illustrate the relationship between a matrix and its inverse.

Matrix Inverse

For a matrix A, its inverse A^{-1} satisfies $AA^{-1} = A^{-1}A = I_n$ where $\mathrm{I_n}$ is the identity matrix.
Below, with an example, we illustrate the relationship between a matrix and its inverse.

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.
Not every square matrix has an inverse; but if a matrix does have an inverse, it is unique.